

논문 2010-1-4

지식베이스에 기반한 다언어 문서 검색

Cross-Lingual Text Retrieval Based on a Knowledge Base

최명복*, 조준**

Myeong-Bok Choi, Jun Jo

요약 웹과 같은 일반 영역을 대상으로 문서를 검색할 때 사용자의 질의 구성은 정보검색 효과에 큰 영향을 준다. 본 논문에서는 일반 사용자들이 웹에서 다언어 문서 검색을 효과적으로 수행할 수 있도록 다언어 지식베이스 기반의 지능형 정보검색 방법을 제안한다. 지식베이스로부터 추론된 지식은 사용자의 연상 작용을 도와 질의를 용이하고 정확하게 구성하여 효과적인 다언어 정보검색을 수행할 수 있도록 한다. 본 논문에서는 이러한 지식베이스 기반의 질의 변경 알고리즘을 개발하고 이를 한국어와 영어 웹 문서를 대상으로 실험하였다. 실험 결과 제안된 질의 변경 알고리즘은 다언어 문서 검색에서 지식베이스를 사용하지 않은 경우에 비해 매우 효과적임을 알 수 있었다.

Abstract User query formation highly acts on the effectiveness of information retrieval when we retrieve documents from the general domain as a web. This thesis proposes a intelligent information retrieval method based on a cross-lingual knowledge base to effectively perform a cross-lingual text retrieval from the web. The inferred knowledge from the cross-lingual knowledge base helps user's word association to make up user query easily and exactly for effective cross-lingual text information retrieval. This thesis develops user's query reformation algorithm and experiments it with Korean and English web. Experimental results show that the algorithm based on the proposed knowledge base is much more effective than without knowledge base in the cross-lingual text retrieval.

Key Words : Cross-lingual text retrieval, knowledge base, thesaurus, intelligent information retrieval, query reformation algorithm

1. 서론

다언어 문서검색(Cross-Lingual Text Retrieval)은 사용자가 작성한 질의어의 언어로 되어 있는 문서뿐 아니라 사용자의 질의어의 언어와 다른 언어들로 되어 있는 문서 자료들로부터 원하는 문서 정보를 검색하는 것을 의미한다. 이에 따라 다언어 문서검색은 질의어의 언어와는 다른 언어로 쓰여진 문서들을 검색하기 위해서 질의어의 언어와 검색 대상 문서 언어를 대응시키기 위하여 크게 두 가지 방법을 사용한다. 하나는 질의어를 검색 대상 문서의 언어로 번역하는 질의 번역(Query Translation) 방

식이고 다른 하나는 검색 대상 문서를 질의어의 언어로 번역하는 문서 번역(Document Translation) 방식이다.

문서 번역 방식은 검색 대상 문서를 질의어의 언어로 변환하기 위해 기계 번역 시스템을 사용한다. 문서 번역은 짧은 길이의 질의어 번역보다는 많은 언어적 정보를 이용할 수 있어 정보 검색의 정확도가 높을 수 있지만 검색 대상이 웹인 경우 현재의 기계번역 기술로 매우 방대한 양의 문서 수집과 정확한 번역은 현실적으로 쉽지 않은 작업이다. 특히 구문적으로만 정확하게 번역하는 기계 번역은 다언어 문서 검색에서 효과적이지 않을 수 있는데 이것은 사용자가 작성한 질의어가 종종 문법적으로 잘 맞지 않기 때문이다^[1].

질의 번역 방식은 사용자의 질의어가 어떤 개념들을

*정회원, 강릉원주대학교 컴퓨터공학부

**비회원, 호주 Griffith University

접수일자 2009.12.31, 수정일자 2010.2.8

이용하여 구성되는지에 따라 크게 두 가지로 세분화 된다. 하나는 통제 개념(Controlled Vocabulary)을 이용한 다언어 문서 검색 방법이고 다른 하나는 자유 개념(Free Vocabulary)을 이용한 다언어 문서 검색이다. 통제 개념을 이용한 다언어 문서 검색은 검색 문서들을 미리 정해 놓은 개념들만을 이용하여 수작업으로 색인해 놓은 후 사용자의 질의어도 문서의 색인에 사용된 개념들과 동일한 개념들만으로 표현하여 검색하는 것을 의미한다. 통제 개념을 이용한 방법은 소규모의 제한된 영역에서 좋은 결과를 보이지만 웹과 같은 일반 영역에서는 많은 양의 문서를 통제 개념을 이용하여 수작업으로 색인해야 하는 어려움이 있다.

자유 개념을 사용한 방식은 질의어를 검색하고자 하는 문서의 언어로 변환하는데 활용하는 언어 자원에 따라 대역사전(Bilingual Dictionary)을 활용하는 방법, 코퍼스(Corpus)를 활용하는 방법, 그리고 시소러스(Thesaurus)를 활용한 방법 등으로 분류된다.

본 논문에서는 다언어 시소러스 기반의 질의 변경을 통한 정보검색 기법을 제안한다. 본 논문에서 제안된 다언어 시소러스는 내부 개념들 간의 관계와 관련도를 가지고 있으며 개념들 간의 관계는 동의, 계층, 그리고 연관 관계의 세 가지 관계가 제공된다. 따라서 다양한 관계에 따른 다언어 시소러스의 다양한 지식을 정보검색에 활용할 수 있다. 이에 따라 제안된 다언어 시소러스로부터 추론된 지식은 네이버(Naver), 구글(Google), 야후(Yahoo) 등과 같은 기존의 상용 포털 시스템에서 제공하는 정보 검색 시스템을 이용한 사용자의 질의를 구성하는데 활용하여 일반 정보검색 사용자가 용이하고 효과적인 정보 검색을 수행 할 수 있게 한다.

본 논문의 나머지 구성은 다음과 같다. 2장에서는 기존의 연구에 대해 간단히 요약한다. 3장에서는 다언어 문서 검색을 위한 다언어 시소러스를 정의하고 시소러스 내의 개념들 간의 추론 규칙 및 추론 방법을 살펴본다. 4장에서는 사용자의 질의 변경을 위한 다언어 시소러스로부터 다양한 지식을 추론해 내는 추론 알고리즘을 제시한다. 5장에서는 실험 결과를 살펴보고 결론을 내린다.

II. 시소러스를 이용한 지식 표현

시소러스는 어휘의 통제와 이용자의 이용 중 무엇에

중점을 두느냐에 따라 그 구성이 크게 달라질 수 있다. 시소러스는 보통 노드와 링크로 구성된다. 노드는 개념을, 그리고 링크는 광의어, 협의어, 동의어와 같은 관계에 의해 개념들 간의 상호 의존성을 표현한다.

단일 언어 정보검색 시스템의 성능을 향상시키기 위해 몇몇 시소러스 기반의 질의 평가 함수들이 제안되었다^[2-5]. 논문^[2]에서 시소러스의 구조는 계층적 개념 그래프(HCG; Hierarchical Conceptual Graph)로 표현된다. HCG는 가중치가 부여된 계층적 시소러스로 단일의 루트 노드(Root Node)를 갖는다. HCG에서 노드들은 색인어들의 집합 T를 구성하며, 링크는 색인어들 간의 "generalization" 관계 G의 정도를 반영하는 가중치 w_{ij} 를 갖는다. 기호 $t_i \in G t_j$ 에서, 색인어 t_i 는 색인어 t_j 보다 좀더 일반적인 광의적 의미의 개념이며, 역으로 t_j 는 t_i 보다 좀더 특정한 협의적 의미를 갖는 개념이다. 다음의 그림 1은 계층적 개념 그래프의 구조를 나타낸다. 또한 논문^[3-5]는 이러한 HCG 구조와 같은 그래프로 구성된 시소러스 기반의 질의 평가 함수들에 대한 연구 논문들이다.

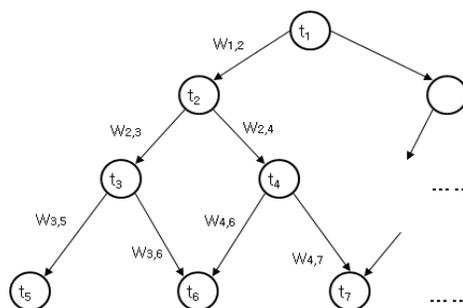


그림 1. 계층적 개념 그래프의 구조
Fig. 1. Structure of HCG

HCG에서 시소러스 내부 개념들 간의 개념적 거리는 두 노드를 연결하는 최단 경로를 찾은 후, 경로상의 링크의 가중치를 합하는 위상적 거리로써 계산된다.

논문^[6]은 퍼지 정보검색을 위한 개념 네트워크를 제안하였다. 개념 네트워크는 노드와 유향 링크(Directed Link)로 구성된다. 노드는 개념 또는 문서를 표현하며, 각 링크는 두 개념들을 연결하든지 또는 특정한 개념과 문서를 연결한다. 또한 링크에는 [0, 1] 사이의 가중치가 부여된다. 링크에 부여된 가중치 w_{ij} 는 개념간의 또는 개념과 문서 사이의 관련 정도를 의미하게 된다. 두 개념들 간의 개념적 거리는 두 노드를 연결하는 최단 경로를 찾은 후, 경로상의 링크의 가중치를 합함으로써 계산된다.

논문^[7]은 논문^[6]에서 제시한 그래프 형태의 개념 네트워크를 모델링하기 위하여 개념 매트릭스를 정의하여 개념 네트워크에서의 그래프 탐색에 따른 시간적 효율성을 해결하였다.

몇몇 다언어 시소러스 기반의 정보검색 기법들^[8-9]이 제안되었다. 논문^[8]은 개념 네트워크^[6]와 동일한 구조의 시소러스 기반의 다언어 문서 검색 방법을 제안하였다. 이 시소러스에서 노드는 용어를 링크는 노드들 간의 부분적 의미 관계를 표현한다. 논문^[9]는 병렬 코퍼스로부터 추출된 용어들을 노드로 하고 이들 노드들을 가중치에 의한 연결한 연관 제약 네트워크(Associate Constraint Network)로 구성된 다언어 시소러스 기반의 정보검색 기법을 연구하였다.

이상의 논문^[2-9]에서 제안된 방법에서는 시소러스를 구성하는 노드들 간의 개념적 거리(Conceptual Distance)를 계산하기 위하여 광의어 또는 협의어와 같은 계층관계만을 고려하든지 또는 개념들의 관계를 고려하지 않고 단지 개념 간의 관련도만을 고려하여 계산한다. 이것은 국제 표준인 ISO2788^[10], ISO5964^[11] 등에서 제안된 개념들 간의 다양한 관계를 표현하지 못함으로써 실제 일반 사용자들이 정보검색에서 다양한 지식을 활용할 수 없게 되는 제한점을 가지게 된다. 또한 이와 같은 연구들^[2-9]에서 사용된 시소러스는 대부분 어휘의 통제 목적에 중점을 두고 있으며 웹 문서와 같은 일반 영역을 대상으로 시소러스를 이용하는 일반 이용자의 측면에 중점을 두고 있지 않다.

웹과 같은 일반 영역에서 정보검색을 하는 사용자는 자국어로 된 탐색어를 사용하여 질의를 구성할 때 찾고자 하는 문서와 관련 있다고 생각하고 있는 의미 있는 연관된 탐색어들을 모두 생각해 내는 것은 종종 어려울 수 있다. 더욱이 다른 언어로 되어 있는 정보까지 얻어내기 위해 질의를 구성할 때 연상되는 탐색어를 모두 생각해 내는 일은 더욱 어려울 수 있다.

본 논문에서는 이러한 제한점들을 개선하기 위해 시소러스 내부 개념들 간의 관계가 ISO2788^[10], ISO 5964^[11] 등에서 제안된 3가지 관계인 동의, 계층, 연관 관계로 구성된 다언어 시소러스 기반의 지능형 정보검색 방법을 제안한다. 따라서 다양한 관계에 따른 다언어 시소러스로부터 추론된 지식을 정보검색에 다양하게 활용할 수 있다. 본 논문에서는 제안된 다언어 시소러스로부터 추론된 지식을 웹 검색에서 사용자의 연상 작용을 도와 질

의를 용이하고 정확하게 구성하여 효과적인 다언어 정보 검색을 수행할 수 있도록 활용한다.

III. 시소러스 내부 개념들 간의 추론 규칙과 추론 기법

사용자의 질의가 주어질 때 그 질의에 사용된 탐색어와 관련된 개념들을 시소러스라는 지식베이스에서 찾아 사용자의 질의를 수정 또는 변경한 후 다시 검색을 수행할 수 있다. 이 때 사용자가 제시한 질의어의 탐색어와 시소러스의 내부 개념들 간의 관련도를 추론하는 방법이 필요하다. 이 부분에서는 시소러스 개념들 간의 관계에 따라 관련도를 추론하는 방법을 설명한다.

1. 추론 규칙

본 논문에서 제안하는 지식 구조는 2가지로 분류된다. 하나는 시소러스이며 다른 하나는 시소러스의 개념들을 연결시키는 개념 매핑 테이블이다. 제안하는 시소러스는 그림 1과 비슷한 형태의 개념 그래프로 구성된다. 그러나 그림 1의 계층적 개념 그래프와는 시소러스가 유지하게 되는 정보와 구성에 있어 다르다. 다시 말해서 본 논문에서 제안하는 시소러스는 그림 2와 같이 노드와 무방향 링크(Undirected Link)로 구성되며 시소러스 내부 개념들 간의 관계가 3가지 관계로 분류되어 표현된다. 그림 2에서 C_i 는 시소러스의 개념이며, D_{ij} 는 개념들 간의 관련도, 그리고 S 는 개념들 간의 3가지 관계들 중의 하나를 나타낸다.

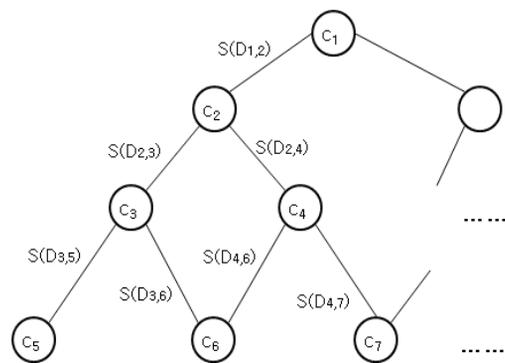


그림 2. 개념 그래프의 구조
Fig. 2. Structure of Conceptual Graph

[규칙3]

C_1	R	C_2	$(D_{1,2})$	자동차	R	highway	$(D_{1,2})$
C_2	R	C_3	$(D_{2,3})$	highway	R	speed camera	$(D_{2,3})$
C_1	R	C_3	$(D_{1,3}(R))$	자동차	R	speed camera	$(D_{1,3}(R))$

[규칙3의 예]

여기서 C_1, C_2, C_3 는 개념이며, $D_{1,2}, D_{2,3}$ 은 개념들 간의 관련도이다(단, $D_{ij} \in [0, 1]$). H는 계층관계, E는 동의 관계, 그리고 R은 연관관계를 의미한다. $D_{1,3}(H), D_{1,3}(E), D_{1,3}(R)$ 등은 $D_{1,2}, D_{2,3}$ 을 사용한 C_1 과 C_3 간의 추론된 관련도와 관계가 된다. 예로 $D_{1,3}(H)$ 는 개념 t_1 과 t_3 사이의 관련도의 계산 결과가 $D_{1,3}$ 이고 H(계층관계)로 추론됨을 의미한다. 제안된 추론규칙 오른쪽은 그 추론규칙을 적용한 하나의 예를 보여준다. [규칙1의 예]는 "plant은 생물이고 flower은 plant이므로 flower은 생물이다"라고 말할 수 있다. 이러한 추론규칙들은 다양한 예제를 통한 관찰에 바탕을 두고 있으며 인식론적으로 비교적 잘 받아들여지는 규칙들이다. 비슷한 방법으로 다음의 [규칙4]에서 [규칙9]까지를 생각할 수 있다. [규칙4]부터 [규칙9]까지에 각각 표기되어 있는 $(D_{1,3}(H_E)), (D_{1,3}(E_H)), (D_{1,3}(H_R)), (D_{1,3}(R_H)), (D_{1,3}(E_R))$, 그리고 $(D_{1,3}(R_E))$ 는 각각의 규칙에서 추론된 관련도와 관계가 된다. 예를 들어 [규칙4]에서 $(D_{1,3}(H_E))$ 는 계층관계와 동의관계로 구성되어 있는 경우 개념 C_1 과 C_3 사이의 관련도는 $D_{1,3}$ 으로 계산되며 대문자 H인 계층관계로 추론됨을 의미한다. 또한 이러한 규칙들은 단언어가 아닌 단일 언어로만 되어 있는 시소러스 내부 개념들 간의 관계에도 그대로 적용되는 규칙들이다.

[규칙4]

C_1	H	C_2	$(D_{1,2})$	C_1	E	C_2	$(D_{1,2})$
C_2	E	C_3	$(D_{2,3})$	C_2	H	C_3	$(D_{2,3})$
C_1	H	C_3	$(D_{1,3}(H_E))$	C_1	H	C_3	$(D_{1,3}(E_H))$

[규칙5]

C_1	E	C_2	$(D_{1,2})$	C_1	H	C_2	$(D_{1,2})$
C_2	R	C_3	$(D_{2,3})$	C_2	R	C_3	$(D_{2,3})$
C_1	R	C_3	$(D_{1,3}(E_R))$	C_1	R	C_3	$(D_{1,3}(H_R))$

[규칙6]

C_1	E	C_2	$(D_{1,2})$	C_1	H	C_2	$(D_{1,2})$
C_2	R	C_3	$(D_{2,3})$	C_2	R	C_3	$(D_{2,3})$
C_1	R	C_3	$(D_{1,3}(E_R))$	C_1	R	C_3	$(D_{1,3}(H_R))$

[규칙7]

C_1	H	C_2	$(D_{1,2})$	C_1	E	C_2	$(D_{1,2})$
C_2	R	C_3	$(D_{2,3})$	C_2	R	C_3	$(D_{2,3})$
C_1	R	C_3	$(D_{1,3}(H_R))$	C_1	R	C_3	$(D_{1,3}(E_R))$

[규칙8]

C_1	R	C_2	$(D_{1,2})$	C_1	R	C_2	$(D_{1,2})$
C_2	H	C_3	$(D_{2,3})$	C_2	E	C_3	$(D_{2,3})$
C_1	R	C_3	$(D_{1,3}(R_H))$	C_1	R	C_3	$(D_{1,3}(R_E))$

[규칙9]

C_1	R	C_2	$(D_{1,2})$	C_1	R	C_2	$(D_{1,2})$
C_2	H	C_3	$(D_{2,3})$	C_2	E	C_3	$(D_{2,3})$
C_1	R	C_3	$(D_{1,3}(R_H))$	C_1	R	C_3	$(D_{1,3}(R_E))$

2. 개념들 간의 관련도 계산을 위한 추론 기법

여기에서는 앞에서 언급한 추론 규칙을 이용하여 실제적으로 개념들 간의 관련도를 계산하는 인식론적 추론 기법을 설명하려고 한다. 앞 절의 규칙들에서 2가지의 사실을 발견할 수 있다. 첫째 동의 관계로만 연결된 규칙보다는 계층 및 연관관계에 의해서 연결된 규칙들이 더 많이 전이에 의해 영향을 받는다. 왜냐하면 계층관계와 연관관계는 동의관계 보다는 의미상의 거리가 일반적으로 더 멀기 때문이다. 둘째 계층, 동의 및 연관관계에서 서로 같은 관계들 간의 추론 결과는 링크의 개수에 의존한다. 즉 링크의 개수가 많아질수록 추론 결과의 의미적 밀접도는 점점 더 약해진다. 또한 앞 절에서 설명된 것과 같이 본 논문에서 제안하는 지식베이스의 구조는 개념들 간의 의미적 관계들로 구성된 정보 구조로 식 (1)과 같은 퍼지 개념 관계로 표현된다. 이러한 2가지의 조건을 만족시키고 퍼지 개념을 반영하기 위해 본 논문에서는 개념들이 같은 관계로만 연결된 규칙의 추론에 다음과 같은 퍼지 관계 연산 식 (2) - (4)를 적용할 것이다.

$$C_i = C_k \text{ 경우,} \\ D_{i,k}(E)=D_{i,k}(H)=D_{i,k}(R)=1 \quad (2)$$

$$C_i \neq C_k \text{ 경우,} \\ \mu_R(x, z) = \text{Max Min}[\mu_R(x, y), \mu_R(y, z)] \\ y \in Y$$

$$\text{(동의관계인 경우)}(\text{단, } (x, z) \in X^2) \quad (3)$$

$$\mu_R(x, z) = \text{Max} [\mu_R(x, y) \cdot \mu_R(y, z)] \\ y \in Y$$

$$\text{(계층, 연관관계인 경우)}(\text{단, } (x, z) \in X^2) \quad (4)$$

또한 서로 다른 관계로 개념들이 연결된 경우에는 다음과 같은 추론 수식 (5) - (8)를 사용할 것이다.

$$D_{i,k}(E_R) = (D_{i,k}(E) + D_{i,k}(R)) / 2 \quad (5)$$

$$D_{i,k}(R_E) = (D_{i,k}(R) + D_{i,k}(E)) / 2 \quad (6)$$

$$D_{i,k}(H_R) = (D_{i,k}(H) + D_{i,k}(R)) / 2 \quad (7)$$

$$D_{i,k}(R_H) = (D_{i,k}(R) + D_{i,k}(H)) / 2 \quad (8)$$

IV. 사용자의 질의 변경

이 부분에서는 사용자가 필요한 문서를 검색하기 위해 질의를 주었을 때 사용자의 질의를 변경할 수 있는 질의 변경 알고리즘을 설명한다. 질의 변경 알고리즘에 의해 생성되어 제시되는 개념들은 사용자의 연상 작용을 도와 사용자의 질의를 좀 더 용이하고 정확하게 구성함으로써 다언어 문서를 효과적으로 검색할 수 있게 한다.

1. 질의 변경 시스템

본 논문에서 제안하는 질의 변경 시스템은 그림 3과 같은 구조를 가진다. 처음에 사용자의 질의가 주어지면 질의 분석기는 사용자의 질의를 분석한다. 분석된 질의는 연상 개념 추론기에 넘겨진다.

연상 개념 추론기에서는 사용자의 질의어에 있는 탐색어와 관련된 개념들을 포함하고 있는 개념들을 지식베이스인 다언어 시소러스를 탐색하여 추론해 낸다. 찾아진 개념들은 사용자 질의어에 있는 모든 탐색어에 관련된 정도에 따라 관련도와 함께 순서화 되어 사용자에게 제시된다. 순위화된 개념들은 사용자의 질의에 있는 탐색어와 관련된 연상 개념들로 볼 수 있다. 사용자는 추론된 연상 개념들 중 필요한 개념들을 이용하여 질의를 변경한 후 검색할 수 있게 된다. 즉 이와 같이 변경된 질의는 기존의 포탈인 네이버(Naver), 구글(Google), 야후(Yahoo) 등과 같은 웹 포털 정보 검색 시스템에 넘겨져 정보 검색을 수행할 수 있게 된다.

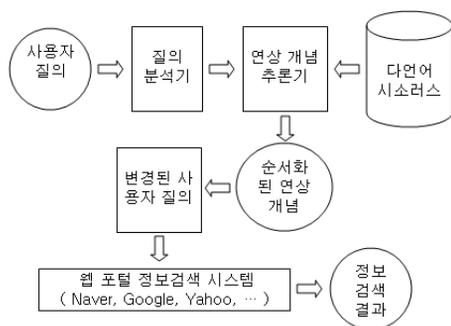


그림 3. 질의 변경 시스템
Fig. 3. Query Reformulation System

여기서 질의 변경은 자동으로도 수행 될 수도 있다. 자동으로 질의를 변경하는 방법은 연상 개념 추론기의 결과인 순위화된 연상 개념들 중에서 사용자의 질의의 탐

색어와 관련도가 일정 값 이상인 개념들만을 사용하여 질의를 확장 한 후 정보 검색을 수행하면 된다.

2. 질의 변경 알고리즘

그러면 이제 연상 개념 추론기에서 다언어 시소러스를 이용하여 사용자의 질의에 있는 탐색어와 관련된 개념을 내포하고 있는 개념들을 탐색하여 추론해 내는 과정을 살펴보자.

사용자의 질의에 있는 탐색어는 그 탐색어의 중요도를 나타내는 값과 함께 구성될 수 있다. 하지만 보통 웹 등과 같은 대부분의 정보검색 시스템에서는 이러한 탐색어의 중요도를 허용하지 않는다. 또한 웹을 대상으로 한 많은 사용자들은 정보검색을 위해 사용자의 질의를 구성할 때 사용자가 찾고자 하는 개념을 내포하고 있는 키워드들 몇 개만을 사용하여 질의를 구성하는 경우가 대부분이다. 이러한 분석에 의해 본 논문에서는 사용자의 질의가 다음의 식 (9)과 같이 구성된다고 가정할 것이다. 즉 식 (9)은 탐색어의 중요도가 명시되지 않은 질의어로 구성되어 있다. 이것은 탐색어로 제시된 중요도가 모두 1로 여길 수 있다. 여기서 C_1, C_2, \dots, C_n 은 탐색어들이다. 이 탐색어들은 사용자의 인지 구조 속에 가지고 있는 검색하고자 하는 문서를 대표하는 개념들이라고 볼 수 있다.

$$Q(C_1, C_2, \dots, C_n) \tag{9}$$

식 (9)과 같은 사용자의 질의가 주어질 때 다음의 알고리즘은 다언어로 되어 있는 각 언어별 시소러스로부터 사용자가 제시한 질의의 탐색어와 관련된 개념들을 관련 정도에 따라 순위화된 목록으로 추론한 후 이를 이용하여 새로운 질의를 구성하는 질의 변경 알고리즘이다. 여기서 다언어 시소러스는 두 개의 서로 다른 언어 L_1 과 L_2 로 구성되어 있고 사용자의 질의는 L_1 언어를 사용하여 구성될 때, 사용자의 질의에 관련된 L_1 언어로 되어 있는 문서와 L_2 언어로 되어 있는 문서를 검색하기 위한 질의 변경 방법을 살펴본다.

- (1) 사용자 질의에 있는 탐색어들과 밀접하게 관련된 개념들을 L_1 시소러스로부터 찾는다.
- (1-1) 다언어의 L_1 시소러스로부터 사용자의 질의에 있는 탐색어 C_1 을 찾는다. 찾아진 개념을 루트 개념(RC1; Root Concept 1)이라고 부

- 를 것이다.
- (1-2) RC1와 밀접하게 관련되어 있는 L_1 시소러스 내의 개념들만을 선택하기 위해 RC1를 중심으로 경계 값(BV; Bound Value)을 정한다. 여기서 RC1에 밀접하게 관련되어 있는 개념들의 기준은 다양할 수 있다. 본 알고리즘에서는 BV를 개념 RC1로부터의 위상적 거리(Topological Distance)에 기반하여 시스템 구현에 따라 조정할 수 있도록 한다. 예를 들어 개념 RC1로부터 위상적 거리 5에 해당되는 개념들만을 RC1과 밀접하게 관련된 개념으로 간주할 수 있다. 이와 같은 위상적 거리에 기반한 BV는 앞의 3장에서 설명된 수식(2)-(8)의 의미를 적절하게 반영하게 된다. 여기서 BV 이내에 해당되는 L_1 시소러스 내의 개념들을 L_1 -연상 개념 집합이라고 부를 것이다.
- (1-3) RC1라는 개념(C_1)과 L_1 -연상 개념 집합에 포함된 각각의 개념(C_j)에 대해 관련도를 추론해 낸다. 관련도의 계산은 동의, 계층, 연관 관계에 따라 앞의 수식(2)-(8)를 이용하여 추론한다. 그러면 RC1와 관련된 모든 개념들의 쌍(C_1, C_j)이 관련도와 함께 얻어진다. 얻어진 개념 쌍과 관련도를 관련도 테이블(RDT1; Relational Degree Table 1)에 포함시킨다. 여기서 j 는 BV 이내에 있는 L_1 시소러스 개념들의 개수에 의존한다.
- (1-4) (1-3)에서 구해진 RDT1 내의 개념들을 관련도에 따라 순위화 시킨다.
- (1-5) 사용자의 질의에 있는 C_2, \dots, C_n 에 대해 각각 (1-1)부터 (1-4)를 수행한다. 단, 이 때 (1-1)부터 (1-4)의 C_1 을 각각 C_2, \dots, C_n 으로 대체한 후 수행하면 된다.
- (2) 사용자 질의에 있는 탐색어들과 밀접하게 관련된 개념들을 L_2 시소러스로부터 찾는다.
- (2-1) 사용자의 질의에 있는 탐색어 C_1 을 다언어 개념 매핑 테이블(CLCMT; Cross-Lingual Conceptual Mapping Table)에서 찾는다. 찾아진 개념을 RC2(Root Concept 2)라고 부를 것이다.
- (2-2) RC2에 대응되는 L_2 시소러스의 개념을 CLCMT에서 찾는다. 찾아진 개념을 MC_RC2(Mapping Concept of Root Concept 2)라고 부를 것이다.
- (2-3) L_2 시소러스에서 개념 MC_RC2를 찾는다. 그리고 (1-2)와 마찬가지로 MC_RC2와 밀접하게 관련되어 있는 L_2 시소러스 내의 개념들만을 선택하기 위해 MC_RC2를 중심으로 경계 값(BV; Bound Value)을 정한다. 본 알고리즘에서는 (1-2)와 마찬가지로 BV를 개념 MC_RC2로부터의 위상적 거리(Topological Distance)에 기반하여 시스템 구현에 따라 조정할 수 있도록 한다. 본 알고리즘에서는 개념 MC_RC2로부터 위상적 거리 5에 해당되는 개념들만을 MC_RC2와 밀접하게 관련된 개념으로 간주하였다. 여기서 BV 이내에 해당되는 L_2 시소러스 내의 개념들을 L_2 -연상 개념 집합이라고 부를 것이다.
- (2-4) CLCMT에 있는 RC2와 MC_RC2 개념 사이의 관련도를 참조하여 개념 RC2라는 개념(C_1)과 L_2 -연상 개념 집합에 포함된 각각의 개념(C_j)에 대해 관련도를 추론해 낸다. 관련도의 계산은 동의, 계층, 연관 관계에 따라 수식(2)-(8)를 이용하여 추론한다. 그러면 RC2와 관련된 모든 개념들의 쌍(C_1, C_j)이 관련도와 함께 얻어진다. 얻어진 개념 쌍과 관련도를 관련도 테이블(RDT2; Relational Degree Table 2)에 포함시킨다. 여기서 j 는 BV 이내에 있는 L_2 시소러스 개념들의 개수에 의존한다.
- (2-5) (2-4)에서 구해진 RDT2 내의 개념들을 관련도에 따라 순위화 시킨다.
- (2-6) 사용자의 질의에 있는 C_2, \dots, C_n 에 대해 각각 (2-1)부터 (2-5)를 수행한다. 단, 이 때 (2-1)부터 (2-5)의 C_1 을 각각 C_2, \dots, C_n 으로 대체한 후 수행하면 된다.
- (3) 사용자의 탐색어와 관련된 관련도 테이블 RDT1과 RDT2의 개념들을 사용자에게 제시하면 사용자는 필요한 개념들을 선택하여 질의를 변경하게 된다.

V. 실험 및 결론

본 논문에서 제안된 다언어 시소러스는 시소러스 내부 개념들 간에, 그리고 두 언어로 구성된 각각의 시소러스 내부 개념들 간에 관련도를 [0, 1] 사이의 퍼지 값으로 표현하였다. 이에 따라 시소러스 내부 개념들 간의 추론 방법은 퍼지 이론에 기반하며 다분히 인식론적이다. 이와 같은 구조의 시소러스인 지식베이스를 이용한 정보검색의 목적은 인간 전문가를 시뮬레이션 하는 것이다^[4]. 이와 같은 맥락에 의해 본 논문에서는 3명의 컴퓨터 전문가와 30명의 대학생들을 이용하여 본 논문에서 제안된 알고리즘에 대한 실험을 수행하였다. 실험을 위해 현재 실제로 일어나고 있는 다음과 같은 컴퓨터 환경을 문제로 제시하고 이를 정보검색을 통해 해결하는 실험을 수행하였다.

(제시된 문제)

“컴퓨터의 전원 버튼을 눌러 컴퓨터를 켜올 때 컴퓨터 화면에 아무런 내용도 표시되지 않는 경우 이를 해결하기 위한 방법은 무엇인가? 인터넷의 웹에서 검색해보아라. 단 현재 이 컴퓨터는 윈도우 xp가 설치되어 있다.”

실험을 위해 우선 컴퓨터 전문가 3인에 의해 이 문제에 대한 적절한 답을 가지고 있는 웹 문서를 선별하였다. 한국어를 위해 네이버에서 20개의 적합한 문서가 그리고 영어를 위해 구글에서 65개의 적합한 문서가 선별되었다. 그런 다음 전문가에 의해 선별된 문서들이 실제 일반 정보검색의 사용자들인 대학생 30명에 의해 검색했을 때 얼마나 많이 우선적으로 검색되는지를 단계적으로 실험하였다. 이를 위해 실제로 1대의 컴퓨터에 윈도우 xp를 설치하고 제시된 문제와 동일한 상황을 만들고 30명의 실험 참가 학생들에게 프로젝트를 통해 보여주면서 설명해 주었다.

이와 같은 실험은 첫째 컴퓨터에 대해 전문적인 지식을 크게 갖추고 있지 않은 학생들과 같은 일반 사용자들이 어떤 방법으로 정보검색을 수행하는지를 살펴보고 둘째 본 논문에서 제안된 다언어 시소러스 기반의 알고리즘에 의해 추론된 지식이 사용자들이 단계적으로 이용했을 때 정보검색의 효과가 있는지를 확인하기 위함이다.

시소러스의 구성은 시소러스를 구성하는 전문가들에 의해 다양하게 만들어 질 수 있다. 본 논문에서는 프린스

턴 대학에서 인지과학적 관점에서 만들어진 WordNet^[13]이라는 범용 사전을 참조하여 제시된 문제를 위한 한국어 시소러스 및 영어 시소러스를 컴퓨터 전문가 3인에 의해 만들었다. 그림 4는 한국어 시소러스의 예를 보여준다.

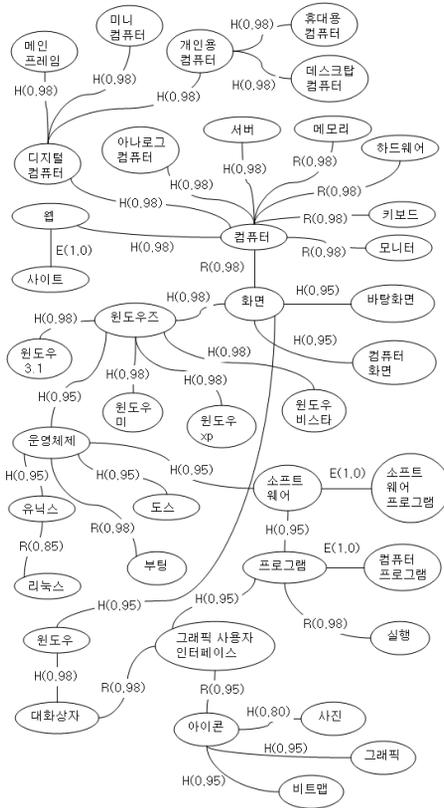


그림 4. 실험용 예제 시소러스
Fig. 4. Sample Thesaurus for Experimentation.

그림에서 E, H, R은 시소러스에서 개념들 간의 관계가 각각 동의, 계층, 연관 관계로 연결되어 있음을 나타낸다. 또한 괄호안의 숫자들은 [0.0, 1.0] 사이의 퍼지 값으로 개념들 간의 관련 정도를 나타낸다. 이 관계와 관련도는 컴퓨터 전문가 3인에 의해 부여 되었으며 특히 관련도는 전문가 3인이 제시한 관련도 값 중에서 가장 작은 값으로 부여하였다.

영어 시소러스도 동일한 한국어 시소러스와 동일한 구조와 관계, 관련도가 부여되었으며 한국어 시소러스와 영어 시소러스의 개념을 각각 1:1로 연결시키는 다언어 개념 매핑 테이블은 다음의 표 2와 같이 구성하였다. 이 실험에서는 표 2와 같이 한국어와 영어 시소러스 내의 개

념들 간의 관련도 $D_{ij}(C_i, C_j)$ 는 모두 1.0으로 부여하였다.

표 2. 한국어-영어에 대한 다언어 개념 매핑 테이블
Table 2. Cross-Lingual Concept Mapping Table of Korean-English

한국어 시소러스 내의 개념	영어 시소러스 내의 개념	$D_{ij}(C_i, C_j)$ $\in [0, 1]$
컴퓨터	computer	1.0
윈도우즈	windows	1.0
.	.	.
아이콘	icon	1.0

실험은 3단계로 분류하여 수행되었다. 첫 번째 단계는 시소러스의 추론된 연상 개념이 제공되지 않은 상태에서 수행되었고 두 번째와 세 번째 단계는 시소러스의 추론된 연상 개념이 제공된 상태에서 정보검색이 행해졌다. 실험은 좀 더 정확한 실험을 위해 한국어와 영어로 분류하여 진행하였으며 실험을 위한 검색엔진은 사용자들에게 친숙한 네이버와 구글 검색엔진을 이용하였다.

실제로 정보검색을 수행할 때 대부분의 사용자들은 사용자의 질의에 관련된 것으로 검색된 상위 웹 페이지들 중에서 잘해야 3~4페이지 정도만을 읽게 된다. 즉 정보검색에서 5페이지 이후의 문서들은 관련된 문서로 검색되어 제시되었다 해도 대부분 무시해 버리는 경향이 있다. 또한 이러한 검색 결과에서 원하는 결과가 없는 경우 사용자들은 질의의 탐색어의 변경, 추가 및 삭제 작업을 거쳐 새로운 정보검색을 수행하게 된다. 이러한 관점을 반영하여 본 논문에서는 네이버와 구글 검색엔진에서 검색된 상위 결과로부터 각각 100, 500개의 문서들로 제한하여 실험하였다.

첫 번째 단계에서는 앞에서 제시된 문제를 대학생 30명에게 충분히 설명한 후 시소러스의 추론된 연상 개념이 제공되지 않은 상태에서 한글 문서를 찾기 위해 네이버에서 정보검색을 수행하였다. 네이버에서 검색한 결과의 페이지들 중에서 우선적으로 제시된 상위 페이지들의 문서 1부터 10페이지(총 문서 수는 100개)에서 제시된 문제에 대한 적절한 답으로 전문가 3인에 의해 선별된 문서와 동일한 문서가 얼마나 많이 검색 되었는지를 알아보았다. 이에 대한 검색 결과는 다음 표 3에서 1단계, 한국어 검색어 부분에 나타나 있다. 표 3에서 기호 "&"는 AND 검색을 의미한다. 이 검색 결과에서 보는 것과 같

이 30명의 학생들은 정보검색을 위해 키워드를 선정하는데 7개의 그룹으로 분류되어 검색하였으며 주로 키워드는 1개 내지는 2개만을 사용하여 검색하였음을 알 수 있다. 표의 네이버 검색 결과에 있는 것과 같이 키워드를 "컴퓨터 화면", "윈도우 xp", "컴퓨터 & 부팅"으로 검색한 사용자는 적절한 문서를 하나도 검색하지 못했으며 "바탕화면" 또는 "컴퓨터 & 실행"으로 검색한 사용자는 각각 1개의 적절한 문서를, 그리고 "아이콘" 또는 "윈도우 & 아이콘"으로 검색한 사용자는 각각 2개의 적절한 문서를 검색하였음을 알 수 있다. 즉 1단계에서는 사용자들이 적절한 문서를 많이 검색하지 못했음을 알 수 있다.

계속해서 사용자들은 네이버에서 사용한 키워드들과 동일한 의미의 영어 키워드들을 사용하여 구글에서 검색을 수행하였다. 이 과정에서 사용된 정확한 영어 키워드들은 좀 더 정확한 실험을 위해 전문가에 의해 사용자들에게 알려주었다. 구글에서 검색한 결과의 페이지들 중에서 우선적으로 제시된 상위 페이지들의 문서 1부터 5페이지(총 문서 수는 500개)에서 제시된 문제에 대한 적절한 답으로 전문가 3인에 의해 선별된 문서와 동일한 문서가 얼마나 많이 검색 되었는지를 알아보았다. 이에 대한 검색 결과는 표 3에서 1단계, 영어 검색어 부분에 나타나 있다. 이 검색 결과에서 보는 것과 같이 키워드 "window & icon"을 사용하여 검색한 그룹의 경우 1개의 적절한 문서만이 검색되었고 나머지 그 외의 키워드를 사용한 그룹의 경우에는 상위 500개 문서에서 적절한 문서가 전혀 검색되지 않았음을 알 수 있다.

두 번째 단계에서는 우선 그림 4와 같은 한국어 및 영어 시소러스와 다언어 개념 매핑 테이블로부터 IV장에서 제안한 질의 변경 알고리즘을 이용하여 첫 번째 단계의 각 사용자 질의의 탐색어와 관련된 다언어 시소러스 개념들에 대해 관련도를 추론하였다. 결과로 나온 관련도의 값에 따라 순위화된 시소러스의 연상 개념 리스트가 각 사용자에게 주어졌다. 이를 위해 본 연구에서는 그림 3의 질의 변경 시스템을 C언어를 이용하여 간단히 구현하였다.

예로 1단계의 검색에서 "아이콘"으로 질의를 작성한 그룹에게는 다음의 표 4와 같은 연상 개념 리스트가 제공되었다. 다시 말해서 1단계 검색에서 사용한 키워드의 그룹에 따라 각각 다른 연상 개념 리스트가 제공되었다. 제공된 연상 개념 리스트를 이용하여 사용자들은 자신의 질의를 수정한 후 네이버에서 다시 검색을 시도하였다.

그 검색 결과가 표의 2단계, 한국어 검색어 부분에 나타나 있다. 비슷하게 1단계에서 사용한 영어 탐색어에 관련된 연상 개념 리스트도 각각의 그룹에 제공되었고 이를 이용하여 질의를 한국어와 동일하게 재구성한 후 구글에서 다시 검색을 시도하였다. 검색 결과가 표의 2단계, 영어 검색어 부분에 나타나 있다.

표 3. 단계별 사용자의 질의에 대한 검색 결과
Table 3. Retrieval Results of User Query with a Level.

한국어 검색어	네이버	영어 검색어	구글	
컴퓨터&화면	0	computer&screen	0	1 단계
바탕화면	1	desktop	0	
아이콘	2	icon	0	
윈도우 xp	0	window xp	0	
컴퓨터&부팅	0	computer&booting	0	
컴퓨터&실행	1	computer&execution	0	
윈도우&아이콘	2	window&icon	1	2 단계
컴퓨터&바탕화면	4	computer&desktop	2	
바탕화면&아이콘	4	desktop&icon	3	
윈도우 xp&바탕화면	4	window xp&desktop	5	
부팅&바탕화면	4	booting&desktop	3	
윈도우 xp&아이콘	5	window xp&icon	4	
컴퓨터&실행&바탕화면	3	computer&execution&desktop	2	
윈도우 xp&부팅&바탕화면	3	w i n d o w xp&booting&desktop	4	
컴퓨터&부팅&바탕화면	6	computer&booting&desktop	5	
윈도우 xp&바탕화면&아이콘	7	w i n d o w xp&desktop&icon	4	
컴퓨터&바탕화면&아이콘	8	computer&desktop&icon	11	3 단계
컴퓨터&바탕화면&아이콘&윈도우 xp	10	computer&desktop&icon&window xp	16	
컴퓨터&실행&바탕화면&아이콘	11	computer&execution&desktop&icon	15	
컴퓨터&실행&바탕화면&아이콘&윈도우 xp	13	computer&execution&desktop&icon&window xp	21	

2단계 검색 결과는 한국어 및 영어 문서 검색에서 모두 1단계의 검색에 비해 적절한 문서들이 좀 더 많이 검색되었음을 알 수 있다.

3단계에서도 2단계와 비슷한 방법으로 2단계의 검색에서 사용한 사용자 질의의 그룹에 따라 각각 다른 연상

개념 리스트가 지식베이스로부터 추론되어 한국어와 영어로 제공되었다. 제공된 연상 개념 리스트를 이용하여 사용자들은 자신의 질의를 수정한 후 다시 검색을 시도하였다.

한국어 및 영어 문서의 검색 결과가 표 3의 3단계, 한국어 검색어 부분과 3단계, 영어부분에 나타나 있다. 3단계에서 한국어 문서 및 영어 문서가 모두 1, 2단계의 검색에 비해 더 많이 검색되었음을 알 수 있다.

본 실험의 결과로 웹의 일반 사용자들의 정보 검색 행동은 몇 가지 특징을 가지고 있음을 알 수 있다. 첫째는 검색 결과에서 보는 것과 같이 단계별로 비교해 볼 때 적합한 문서의 검색에 있어서 한국어 문서와 영어 문서가 비슷하게 검색되었음을 알 수 있다. 이것은 한국어를 주로 사용하는 이용자와 영어를 주로 사용하는 이용자들이 거의 비슷한 개념들을 선택하여 웹 문서를 생성하고 사용한다는 것을 의미한다. 이러한 의미는 다언어로 된 문서를 검색할 때 다언어 시소러스의 구성과 검색 알고리즘을 개발할 때 좋은 분석 자료가 될 수 있다. 둘째는 일반 웹의 사용자들이 정보검색을 수행 할 때 우선 본인의 머릿속에 가장 크게 관련 있다고 떠오르는 개념 하나 또는 두 개 정도를 탐색어로 사용하여 검색을 수행한다. 검색엔진에 의해 검색되어 우선적으로 제시된 문서들을 읽어가며 자신이 찾는 정보가 있는지를 살펴본다. 이 과정에서 본인이 미처 생각하지 못했던 개념들이 나오면 이를 이용하여 새로운 검색을 시도하게 된다.

이러한 과정은 원하는 정보가 검색될 때까지 계속된다. 이러한 정보검색 행동은 사용자가 관련 있다고 생각하는 연상 개념들을 찾는 과정과 이를 이용한 새로운 검색의 시도가 이루어진다고 볼 수 있다. 이러한 과정은 검색의 시간적 측면과 정확한 문서를 검색하는 정확도 측면에 큰 영향을 줄 수 있으며 때에 따라서는 원하는 검색 결과를 얻는데 매우 많은 시간이 소요되거나 적절한 검색 결과를 얻지 못할 수도 있다.

이를 해결하는 하나의 방안으로 본 연구에서는 정보 검색 사용자의 연상 작용을 도와 정보검색을 효과적으로 수행 할 수 있는 다언어 시소러스 기반의 다언어 웹 문서를 검색 할 수 있는 추론 시스템과 알고리즘을 제시하고 이를 이용한 실험을 수행하였다. 앞의 실험 결과를 그래프로 요약하면 다음의 그림 5와 같이 된다. 그림 5는 한국어와 영어 각각에 대한 단계별 검색된 문서의 합계를 기준으로 작성한 것이다. 이 그래프에서 보는 것과 같이

한국어 문서에서는 지식베이스인 시소러스를 사용하지 않은 1단계의 검색에 비해 본 논문에서 제안된 시소러스를 사용한 2단계 및 3단계의 검색에서 각각 24%와 55% 검색 성능이 증가하였으며, 영어의 경우에도 1단계에 비해 지식베이스를 사용한 2단계 및 3단계의 검색에서 각각 23%와 74% 검색 성능이 증가하였음을 알 수 있다.

표 4. 질의어 "아이콘"에 대한 순위화 된 다언어 연상 개념 리스트

Table 4. Cross-Lingual Association Concept List of User Query with "icon"

관련된 개념 (한국어)	관련도	관련된 개념 (영어)	관련도
그래픽	0.95	graphic	0.95
그래픽 사용자 인터페이스	0.95	GUI(Graphic User Interface)	0.95
비트맵	0.95	Bitmap	0.95
대화상자	0.931	photo	0.931
윈도우	0.912	window	0.912
컴퓨터 프로그램	0.903	computer program	0.903
프로그램	0.903	program	0.903
실행	0.885	execution	0.885
화면	0.867	screen	0.867
소프트웨어	0.857	software	0.857
소프트웨어 프로그램	0.857	software program	0.857
윈도우즈	0.85	windows	0.85
컴퓨터	0.85	computer	0.85
운영체제	0.84	operating system	0.84
바탕화면	0.823	desktop	0.823
부팅	0.823	booting	0.823
컴퓨터 화면	0.823	computer screen	0.823
사진	0.8	photo	0.8
도스	0.798	dos	0.798
유닉스	0.798	UNIX	0.798

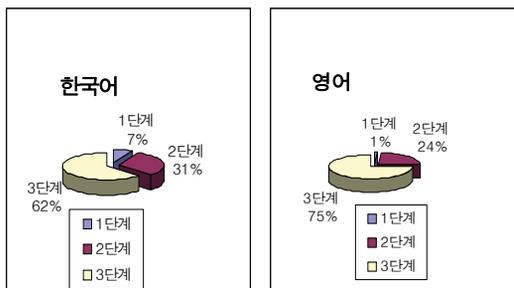


그림 5. 성능 비교 그래프
Figure 5. Effectiveness Comparison Graph.

결국 실험 결과의 분석으로 볼 때 사람의 연상 작용을 돕는 다양한 관계와 관련도를 갖는 다언어 시소러스와 같은 지식베이스의 이용은 다언어 정보검색을 훨씬 더 유연하고 효과적으로 수행 할 수 있음을 알 수 있었다. 현재 본 연구는 WordNet^[13]과 비슷한 대규모의 시소러스를 구축하고 이를 온라인에서 자동으로 이용할 수 있는 시스템을 고려하고 있다.

참고문헌

- [1] Rowena Chau and Chung-Hsing Yeh, Fuzzy Conceptual Indexing for Concept-Based Cross-Lingual Text Retrieval, IEEE INTERNET COMPUTING, pp. 14 ~ 21, 2004.
- [2] Kim, Y. W.; Kim, J. H. A model of knowledge based information retrieval with hierarchical concept graph, Journal of Documentation, 46(2), pp. 113-136, 1990.
- [3] Rada, R.; Mili, H.; Bicknell, E., & Blettner, M. Development and application of a metric on semantic nets, IEEE Transactions on systems, man, cybernetics, Vol. 19, No. 1, pp. 17-36, January, 1989.
- [4] McMath, C. F.; Tamaru, R. S.; Rada, R. A graphical thesaurus-based information retrieval system, International Journal of Man-Machine Studies, 31(2), pp. 121-147, 1989.
- [5] Lee, J. H., Kim, M. H., & Lee, J. H. Ranking documents in thesaurus-based boolean retrieval systems, Information Processing & System. 30(1), pp. 79-91, 1994.
- [6] Lucarella, D.; Morara, R. FIRST: Fuzzy information retrieval system, Journal of Information Sciene, Vol. 17, pp. 81-91, 1991.
- [7] Chen, S. M.; Wang, J. Y, Document retrieval using knowledge-based fuzzy information retrieval techniques, IEEE Transactions on systems, man, cybernetics, Vol. 25, No. 5, pp. 793-803, 1995.
- [8] Rowena Chau and Chung-Hsing Yeh, Crossing

the language barrier using fuzzy logic, Springer-Verlag, pp. 768-773, 2005.

[9] Christopher C. Yang, Chih-Ping Wei, K.W. Li. Cross-lingual for multilingual knowledge management, Decision Support Systems, Vol. 45, pp. 596-605, 2008.

[10] ISO 2788, International Organization for Standardization: Guidelines for the Establishment and Development of Monolingual Thesauri, 2nd ed., Geneva, ISO, 1986.

[11] ISO 5964, International Organization for Standardization: Guidelines for the Establishment and Development of Multilingual Thesauri, 2nd ed., Geneva, ISO, 1985.

[12] Zadeh, L. A., Similarity relations and fuzzy orderings, Information Science, Vol 3, pp. 177-206, 1971.

[13] <http://wordnet.princeton.edu/>

※ 이 논문은 2009학년도 강릉원주대학교 장기해외파견 연구지원에 의하여 수행되었음

저자 소개

최 명 복(Myong-Bok Choi)



- 2001년 : 아주대학교 컴퓨터공학과 (박사)
 - 현재 : 강릉원주대학교 컴퓨터공학부 (교수)
 - 2003. 9~현재 한국정보과학회 전산 교육연구회 운영위원
 - 2003. 9~현재 한국정보과학회 전문 대학 학회지·논문지 편집위원
 - 2004. 1~현재 한국컴퓨터산업교육학회 학회지 편집위원
 - 2004. 1~현재 한국인터넷방송통신TV학회 이사
 - 2009년 호주 Griffith University 방문교수
- <주관심분야 : 지능형 정보검색, 퍼지응용, 지식표현, 신경망, 지능형 교통제어, 소프트웨어 공학, 임베디드 응용 시스템>

조 준(Jun Jo)



- 호주 Sydney University(박사)
 - 현재 : 호주 Griffith University 교수
 - FIRA의 SimuroSot 분야 이사
 - Australian Robotics Association (ARA) 회장
 - 2010 국제 로봇 올림피아드 의장
- <주관심분야 : 지능형 로봇 응용 시스템, AI 응용, 지능형 게임 엔진, e-learning, Genetic Algorithms>